

# Variability in the analysis of a single neuroimaging dataset by many teams

Rotem Botvinik-Nezer<sup>1</sup>, Cemal Koba<sup>2,\*</sup> and ... (Full list of authors and affiliations appears in the online version of the paper [1])

<sup>1</sup> *Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA*

<sup>2</sup> *Sano Center for Computational Medicine, Krakow, Poland*

\* [c.koba@sanoscience.org](mailto:c.koba@sanoscience.org)

Data analysis workflows in many scientific domains have become increasingly complex and flexible. Here we assess the effect of this flexibility on the results of functional magnetic resonance imaging by asking 70 independent teams to analyse the same dataset, testing the same 9 ex-ante hypotheses<sup>1</sup>. The flexibility of analytical approaches is exemplified by the fact that no two teams chose identical workflows to analyse the data. This flexibility resulted in sizeable variation in the results of hypothesis tests, even for teams whose statistical maps were highly correlated at intermediate stages of the analysis pipeline. Variation in reported results was related to several aspects of analysis methodology. Notably, a meta-analytical approach that aggregated information across teams yielded a significant consensus in activated regions. Furthermore, prediction markets of researchers in the field revealed an overestimation of the likelihood of significant findings, even by researchers with direct knowledge of the dataset [2,3,4,5]. Our findings show that analytical flexibility can have substantial effects on scientific conclusions, and identify factors that may be related to variability in the analysis of functional magnetic resonance imaging. The results emphasize the importance of validating and sharing complex analysis workflows, and demonstrate the need for performing and reporting multiple analyses of the same data. Potential approaches that could be used to mitigate issues related to analytical variability are discussed.

## References

- [1] Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020). <https://doi.org/10.1038/s41586-020-2314-9>
- [2] Dreber, A. et al. Using prediction markets to estimate the reproducibility of scientific research. *Proc. Natl Acad. Sci. USA* **112**, 15343–15347 (2015).
- [3] Camerer, C. F. et al. Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).
- [4] Camerer, C. F. et al. Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nat. Hum. Behav.* **2**, 637–644 (2018).
- [5] Forsell, E. et al. Predicting replication outcomes in the Many Labs 2 study. *J. Econ. Psychol.* **75**, 102117 (2019).